

\*\*\*\*\*  
 WISEMAN (TM)  
 \*\*\*\*\*

Release 3.1a John F. Collins, BioComputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

Merch\_PP protein - protein database search, using Smith-Waterman algorithm  
 Run on: Sat May 13 08:27:01 2000; Maspar time 19.30 Seconds  
 Tubular output not generated. 644.282 Million cell updates/sec

Title: >US-09-331-631-7  
 Description: (1-525) from US09331631.pep (1 of 3)  
 Perfect Score: 3808  
 Sequence: 1 MWISKSPFIVLIFSLLSFA.....FLNKNONTNVIKTVKASAY 525

Scoring table: PAM 150  
 Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a-geneseq35  
 1:geneseqp

Statistics: Mean 35.694; Variance 169.035; scale 0.211

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description	Pred. No.
1	3808	100.0	525	1 W62831	Theobroma cacao antim	0.00e+00
2	3654	96.0	566	1 R20181	Sequence encoded by 67	0.00e+00
3	1910	50.2	590	1 W62832	Gossypium hirsutum ant	7.10e-157
4	1193	31.3	666	1 W62829	Macadamia integrifolia	1.88e-92
5	1178	30.9	666	1 W62828	Macadamia integrifolia	4.11e-91
6	1149	30.2	625	1 W62830	Macadamia integrifolia	1.60e-88
7	848	22.3	593	1 W62835	zea mays antimicrobia	8.89e-62
8	810	21.3	593	1 W62838	Glycine max antimicrob	2.02e-58
9	796	20.9	614	1 W22149	Peanut allergen Ara hi	3.46e-57
10	796	20.9	614	1 W62834	Arachis hypogaea antim	2.59e-53
11	752	19.7	626	1 W22150	Peanut allergen Ara hi	2.73e-51
12	729	19.1	489	1 W90341	G. max SBP2 protein.	7.49e-51
13	724	19.0	409	1 W90342	G. max truncated SBP2	6.39e-49
14	702	18.4	524	1 W90339	G. max SBP1 protein.	5.41e-47
15	680	17.9	444	1 W90340	G. max truncated SBP1	2.12e-39
16	593	15.6	637	1 W62837	Hordeum vulgare antim	5.11e-04
17	169	4.4	395	1 W03474	Mouse SRV-related prot	2.20e-02
18	147	3.9	186	1 W26536	Trypanosoma cruzi anti	2.61e-02
19	146	3.8	521	1 W74802	Human secreted protein	2.26e-01
20	133	3.5	86	1 W95073	GST-HD fusion protein	2.26e-01
21	133	3.5	86	1 W95078	GST-HD fusion protein	2.26e-01
22	133	3.5	94	1 W95080	GST-HD fusion protein	2.26e-01
23	133	3.5	94	1 W95075	GST-HD fusion protein	2.26e-01

## ALIGNMENTS

ID	Sequence	Score	DB 1:	Length	Mismatches	Indels	Gaps
1	W62831 standard; Protein; 525 AA.	100.0%	DB 1:	525	0	0	0
2	W62831 (first entry)	100.0%	DB 1:	525	0	0	0
3	Theobroma cacao antimicrobial protein.	96.0%	DB 1:	566	0	0	0
4	antimicrobial protein; infestation; control.	50.2%	DB 1:	590	0	0	0
5	Theobroma cacao.	31.3%	DB 1:	666	0	0	0
6	W09827805-A1.	30.9%	DB 1:	666	0	0	0
7	W09827805-A1.	30.2%	DB 1:	625	0	0	0
8	W09827805-A1.	22.3%	DB 1:	593	0	0	0
9	W09827805-A1.	21.3%	DB 1:	593	0	0	0
10	W09827805-A1.	20.9%	DB 1:	614	0	0	0
11	W09827805-A1.	19.7%	DB 1:	626	0	0	0
12	W09827805-A1.	19.1%	DB 1:	489	0	0	0
13	W09827805-A1.	19.0%	DB 1:	409	0	0	0
14	W09827805-A1.	18.4%	DB 1:	524	0	0	0
15	W09827805-A1.	17.9%	DB 1:	444	0	0	0
16	W09827805-A1.	15.6%	DB 1:	637	0	0	0
17	W09827805-A1.	4.4%	DB 1:	395	0	0	0
18	W09827805-A1.	3.9%	DB 1:	186	0	0	0
19	W09827805-A1.	3.8%	DB 1:	521	0	0	0
20	W09827805-A1.	3.5%	DB 1:	86	0	0	0
21	W09827805-A1.	3.5%	DB 1:	86	0	0	0
22	W09827805-A1.	3.5%	DB 1:	94	0	0	0
23	W09827805-A1.	3.5%	DB 1:	94	0	0	0

Qy	301	KLEELLEBQROQKQOQOGGCFRAKREPQIRAIQOQTSPHHNGGERTLAINLSQSPVYS	360
D5b	361	NONGRFEACPEDEFSOEFQNMVAVSAEKLNGALFVEPHYNSKATFVVFTVDGYGTAOMAC	420
Qy	361	NONGRFEACPEDEFSQOFQNMVAVSAEKLNGALFVPHYNSKATFVVFTVDGYGTAOMAC	420
D5b	421	PLHSOSOGSOGSDRRREDEESEEETFGFEGQVKAPLSBGDFVAPAGHAATFFASKD	480
Qy	421	PLHSOSOGSOGSDRRREDEESEEETFGFEGQVKAPLSBGDFVAPAGHAATFFASKD	480
D5b	481	QPLNAVAFGLNAQNNQRIETLGRPFELNHKQNTVIRKETVYKASAY	525
Qy	481	QPLNAVAFGLNAQNNQRIETLGRPFELNHKQNTVIRKETVYKASAY	525

RESULT 2  
 ID R20181 standard: Protein: 566 AA.  
 AC R20181;  
 DT 16-APR-1992 (first entry)  
 DE Sequence encoded by 67 kD T. cacao protein cDNA.  
 KW Cocoa: flavour: vicilin; seed storage protein.  
 OS Theobroma cacao.  
 PN W09119801-A.  
 PD 26-DEC-1991.  
 PE 07-JUN-1991. G00914.  
 PR 11-JUN-1990: GB-013016.  
 PA (MRSC ) MARS UK LTD, Deakin EA, Ashton S;  
 PI Spencer ME, Hodge R, Deakin EA, Ashton S;  
 DR WPI: 92-024418/03.  
 DR N-PSDB: 020377.  
 PT Recombinant cocoa proteins - are responsible for flavour in cocoa  
 beans and produced in large quantities using yeast and bacterial  
 expression vectors  
 PT Claim 4: Fig 2; 59pp: English.  
 PS The inventors claim a 67 kD and 31 kD T. cacao protein, and  
 CC fragments, and encoding DNAs. The 47 kD and 31 kD proteins are  
 CC derived from the 67 kD precursor. T. cacao protein cDNA was  
 CC detected in a cDNA library prepared from immature cocoa beans RNA  
 CC using a probe based on the AA sequence of a CNBR peptide common to  
 CC the 47 kD and 31 kD polypeptides. Homology searches revealed close  
 CC homologues between the 67 kD polypeptide and the vicilins, which are  
 CC seed storage proteins.  
 QO Sequence 566 AA.

Query Match	96.0%	Score 3654	DB 1	Length 566
Best Local Similarity	96.7%	Pred. No. 0.00e+00		
Matches 504	Conservative	10	Mismatches 6	Indels 1
				Gaps 1
Db	1	1	1	1
Qy	1	1	1	1
Db	61	61	61	61
Qy	61	61	61	61
Db	121	121	121	121
Qy	121	121	121	121
Db	181	181	181	181
Qy	181	181	181	181
Db	241	241	241	241
Qy	241	241	241	241
Db	301	301	301	301
Qy	301	301	301	301

Db	361	NONGEEFACEDDESOEQNMVAASAKLNGALEFVHYNSKATFVVYVVDGYGMAAC	420
Qy	361	NONGFEFACEDDESQEQNMVAASAKLNGALEFVHYNSKATFVVYVVDGYGMAAC	420
Db	421	PHLSOSOGSOGSDRREOESESEETFGFEOVKAPLSPGVFVAPACHATFFASKD	480
Qy	421	PHLSOSOGSOGSDRREOESESEETFGFEOVKAPLSPGVFVAPACHATFFASKD	480
Db	481	QPLNVAFGLAQNQNRIFLGLKNLVRQMSSEAKELSPGV	521
Qy	481	QPLNVAFGLAQNQNRIFLGLGRFFLHNQN -INVTKFTV	520

RESULT 3  
ID M62832 standard; Protein; 590 AA.  
AC M62832:  
DT 27-OCT-1998 (first entry)  
DE Gossypium hirsutum antimicrobial protein.  
KW antimicrobial protein; infestation; control.  
OS Gossypium hirsutum.  
PN M09827805-A1.  
PD 02-JUL-1998.  
PT 22-DEC-1997: AU0874.  
PR 20-DEC-1996: AU-004275.  
RA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.  
PI Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP,  
WPI: 98-377279/32.  
PT Novel anti-microbial protein from e.g. *Macadamia integrifolia* -  
useful for controlling microbial infestations of plants or mammals  
PS Claim 1: Page 49-51: 96pp. English  
CC The sequence is that of an antimicrobial protein which can  
be used to control microbial infestations in plants and mammalian  
CC animals.  
SQ Sequence 590 AA:

[illegible]

## RESULT 4

ID	W62829 standard; Protein; 666 AA.
AC	W62829;
DT	27-OCT-1998 (first entry)
DE	Macadamia integrifolia antimicrobial protein.
KW	antimicrobial protein; infestation; control.
OS	Macadamia integrifolia.
FH	Key
FT	Peptide
FT	1..28
FT	/note= "signal peptide"
FT	29..666
FT	/note= "mature protein"
PN	W09827805-A1.
PD	02-JUL-1998.
PF	22-DEC-1997; AU0874.
PR	20-DEC-1996; AU-004275.
PA	(RBR-) COOP RES CENT TROPICAL PLANT PATHOLOG.
PI	Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;
DR	WPI: 98-377279/32.
DR	N-PSDB: V42311.
DR	Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT	useful for controlling microbial infestations of plants or mammals
PS	Claim 1; Page 39-41; 96pp; English.
CC	The sequence is that of an antimicrobial protein which can
CC	be used to control microbial infestations in plants and mammalian
CC	animals.
CC	Sequence 666 AA;
CC	Sequence

[illegible]

FT	Protein	/note="signal peptide"
FT		29, 666
PT		/note="mature protein"
PN	W09827805-A1.	
PD	02-JUL-1998.	
FE	22-DEC-1997; AU0874.	
PR	20-DEC-1996; AU-004275.	
PA	(RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.	
PI	Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;	
DR	WPI: 98-377279/32.	
DR	N-PSDB: V42310.	
PT	Novel anti-microbial protein from e.g. <i>Macadamia integrifolia</i> -	
PT	useful for controlling microbial infestations of plants or mammals	
PS	Claim 1: Page 34-36: 96pp. English	
CC	The sequence is that of an antimicrobial protein which can	
CC	be used to control microbial infestations in plants and mammalian	
CC	animals.	
CC	Sequence	666 AA;
50		

Query Match	30.9%	Score 1178	DB 1	Length 666
Best Local Similarity	40.0%	Pred. No. 4,11e-91		
Matches 191	Conservative 125	Mismatches 139	Indels 23	Gaps 17
Db	137	ETEDRH-OTCOQCEERRRYKKRQKRYEEOREDEEKEVEEEMKEEDKRDPOQREYE	195	
QY	35	ERDPRQYEQEQRCRCEE-ATEEEDQE-QCE-QRCRE-YKEQQRQDE--E-LQRYQ	86	
Db	196	DCRRRC-EQOE-PROQHCOCLREERQROHGRGDDMMNPGRGSGRYEESEEDSDNPY	253	
QY	87	QCQGRCEOOOGOREOQOCCRKCEQYKEQDERGHEHENVHNK-KNRSEEEGQORNPY	145	
Db	254	FDER-SISTPEEHEGHISYLENYGSKILLRAKNRLVLEANPAFVLPHTLADAI	312	
QY	146	FPKRRSQTPREDEEGNFKIDQRAENSPLKGINDRILAMFENPPTFLLPHCDAEAI	205	
Db	313	LIVYGGALKMIHNDRESYVLECGVIRIPATGTYLLINRDNEERLHAKFLQITSTP	372	
QY	206	YFYVINGKGTITFYHKEKESYNORGVVSPAGSTIYVVSQDQKELTAVLALPNSP	265	
Db	373	GOYKEFPAGGONDEPYLSTFSKEILEALNTQTEKLVGFQGOOREGVIIIRASQEQIREL	432	
QY	266	GKVELFPAPANNKPESSYYGAFSEYLETETVNTQREKLELLEDRGKROGGQOGMRKA	325	
Db	433	TRDD--SESHMHIRGEGSSRGDPYLFNKRPVLSNKGAYEVKPEDYRQLODMDSVF	490	
QY	326	KPEIRAIISOQATSPRRHGERLAINLDSOPYSNONGFFPEACEDEFSQFONDVAYS	385	
Db	491	IANTQSSMAGPEFNNSTKVVVVYVASEAVEACPHLSR-HGKRG--GKR-HEE--	544	
QY	386	AFKLNQCAITVPHNSKATVFEVTDGYQAQACPLHSQSGOSGRODRREQEESE	445	
Db	545	EDV--HEZYQARLSKKEALIVLAGHVYVSSGNETLLFAGCINAKNNHNEFLAGR	600	
QY	446	EEITFGEQQKAPLSPEDVYFAPAGNAVTFPAKXDPPLNAVAGELNANQNRITFLAGR	503	
RESULT				
ID	6	W62830 standard; Protein: 625 AA.		
AC		W62830:		
DT	27	OCT-1998 (first entry)		
DE		Macadamia integrifolia antimicrobial protein.		
KW		antimicrobial protein; infestation; control.		
OS		Macadamia integrifolia.		
Key		Location/Qualifiers		
FT	peptide	1..28		
FT		/note="signal peptide"		
FT	Protein	29..666		
FT		/note="mature protein"		
PN	W09827805-A1.			
PD	02-JUL-1998.			
PF	22-DEC-1997.	AU0874.		
PR	20-DEC-1996.	AU-004275.		
RA	(RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.			









QY 187 FEANPNTFLPHHCDAAEALFYVNTNGKGTTFVTYHNKESINYQRTVSVSPAGSTVYVS 246  
DB 213 RDENDKFLAMLHIPVSVSTPGKFEFFAPGGRDESVLSAFSNVWLOAALQTPKSKLEN 272  
QY 247 QNOKELTIAVALALPVN-S-PKYTELFPAGNNKPESTYGAFSYVLEVTYNTQREKLEE 304  
DB 273 VFD-Q--QN--EG--SIFRISR-EQVRLAAPTCKSSWMPFGESKQPFNIFSKRPTISNG 324  
QY 305 ILEEQGQKROOGQOGMFRKAKPEQIRAIISOQATSPRRH-R-GLAINLLSQSPVYSNQ 362  
DB 325 YGRLEVPDDEKSWLQRLNMLFTNTITQSMSTIHYNSATKIALYIDGRHLOJSC 384  
QY 363 NGRFFACPEDEFSQ--FQNMVAVSAFKLNOGALFVPHYNSKATFEVYTDGYAQMAG 420  
DB 385 PHMS--SRSSHS-KHDS-----S---P-S-YHRISSDKPGMVFVPPGHPFVITASNK 431  
QY 421 PHLSRQSGSGSGORRQGRQEESEETFGFQGVAKAPLSPGVFAVAPGHAVTFAASND 480  
DB 432 ENILMICEVYVARNADNKKFTFACK 454  
QY 481 OPLNAVAFGLNMQNNORIFLAGR 503

## RESULT 15

ID W90340 standard; protein: 444 AA.  
AC W90340;  
DT 24-MAY-1999 (first entry)  
DE G. max truncated SBP1 protein.  
KW SBP1; sucrose binding protein; SBP2; sucrose uptake; transgenic plant;  
KW seed; carbohydrate content; soybean.  
OS Glycine max.  
PN W09653086-R1.  
PD 26-NOV-1998.  
PE 21-MAY-1998; U10465.  
PF 22-MAY-1997; US-047568.  
PI (UNITW) UNIV WASHINGTON STATE RES FOUND.  
PI Chao WS, Gimes HD;  
PI WPI: 99-070135/06.  
PT New modified plant sucrose binding proteins - used to develop  
PT transgenic plants which can have enhanced or decreased sucrose  
PT uptake activity in developing seeds  
PS Claim 7, Page 36-37, 58pp; English.  
CC This sequence represents a novel sucrose binding protein, SBP1 isolated  
CC from Glycine max. This protein is used in a method resulting in the  
CC production of a modified plant sucrose binding protein (SBP) which has a  
CC modified amino acid sequence compared to a corresponding wild-type SBP,  
CC and where expression of the modified SBP in a yeast assay system confers  
CC enhanced sucrose uptake compared to the corresponding wild-type SBP.  
CC The products of the invention can be used for producing transgenic plants  
CC which have modified sucrose uptake activity, particularly in developing  
CC seeds. Enhanced sucrose uptake activity in developing seeds may be  
CC desirable where it is an advantage to increase the carbohydrate content  
CC of the seed (e.g. where the seed is the primary plant material harvested,  
CC such as soybean). In contrast, decreased sucrose uptake activity in  
CC seeds might be desirable where the vegetative material of the plant is  
CC harvested. The SBP regulatory regions confer specific or enhanced  
CC expression in developing seeds and so may be used to express any  
CC transgene in developing seeds.  
SQ Sequence 444 AA;

Query Match 17.9%; Score 680; DB 1; Length 444;

Best Local Similarity 33.9%; Pred. No. 5, 41e-47;

Matches 126; Conservative 114; Mismatches 104; Indels 28; Gaps 20;

DB 34 EEEDEPLVCKHCCOCCOQOYTGEDKRVCLQSC-DRYHRMKOERKOIOETREKKEESR 92  
QY 79 EELQROYOCCGRCROEQOQ-GOREOQOCCQKCKWBT-K-EQERG-E-HENYHNK---K 129  
DB 93 EREEOQOEHEODENPYTFEEDKDFETRVETEGGRIRVLKFKTEKSKLQCIENFRLAI 152  
QY 130 NRSEEEGGO--RN--NPYFPRKRSQGTFRDEEGNFKILQRAENSPLPKGINDYRLAM 186  
DB 153 LEARAHTEVSPRHDFSEVVFENIKRAVLGLVSESETEKITLEPGDMIHIPAGTPLYIVN 212

QY 187 FEANPNTFLPHHCDAAEALFYVNTNGKGTTFVTYHNKESINYQRTVSVSPAGSTVYVS 246  
DB 213 RDENDKFLAMLHIPVSVSTPGKFEFFAPGGRDESVLSAFSNVWLOAALQTPKSKLEN 272  
QY 247 QNOKELTIAVALALPVN-S-PKYTELFPAGNNKPESTYGAFSYVLEVTYNTQREKLEE 304  
DB 273 VFD-Q--QN--EG--SIFRISR-EQVRLAAPTCKSSWMPFGESKQPFNIFSKRPTISNG 324  
QY 305 ILEEQGQKROOGQOGMFRKAKPEQIRAIISOQATSPRRH-R-GLAINLLSQSPVYSNQ 362  
DB 325 YGRLEVPDDEKSWLQRLNMLFTNTITQSMSTIHYNSATKIALYIDGRHLOJSC 384  
QY 363 NGRFFACPEDEFSQ--FQNMVAVSAFKLNOGALFVPHYNSKATFEVYTDGYAQMAG 420  
DB 385 PHMS--SRSSHS-KH 396  
QY 421 PHLS-RQSGSG 431

Search completed: Sat May 13 08:27:24 2000  
Job time : 23 secs.